Plant Mol Biol (2006) 61:799–815
DOI 10.1007/s11103-006-0051-6

The 73 kD Subunit of the cleavage and polyadenylation specificity factor (CPSF) complex affects reproductive development in *Arabidopsis*

Ruqiang Xu · Hongwei Zhao · Randy D. Dinkins · Xiaowen Cheng · George Carberry · Qingshun Quinn Li

Received: 14 October 2005 / Accepted: 22 March 2006

Abstract The cleavage and polyadenylation specificity factor (CPSF) is an important multi-subunit component of the mRNA 3′-end processing apparatus in eukaryotes. The *Arabidopsis* genome contains five genes encoding CPSF homologues (AtCPSF160, AtCPSF100, AtCPSF73-I, AtCPSF73-II and AtCPSF30). These CPSF homologues interact with each other in a way that is analogous to the mammalian CPSF complex or their yeast counterparts, and also interact with the *Arabidopsis* poly(A) polymerase (PAP). There are two CPSF73 like proteins (AtCPSF73-I and AtCPSF73-II) that share homology with the 73 kD subunit of the mammalian CPSF complex. AtCPSF73-I appears to correspond to the functionally characterized mammalian CPSF73 and its yeast counterpart. AtCPSF73-II was identified as a novel protein with uncharacterized protein homologues in other multicellular organisms, but not in yeast. Both of the AtCPSF73 proteins are targeted in the nucleus and were found to interact with AtCPSF100. They are also essential since knockout or knockdown mutants are lethal. In addition, the expression level of AtCPSF73-I is critical for *Arabidopsis* development because overexpression of AtCPSF73-I is lethal. Interestingly, transgenic plants carrying an additional copy of the *AtCPSF73-I* gene, that is, the full-length cDNA under the control of its native promoter, appeared normal but were male sterile due to delayed anther dehiscence. In contrast, we previously demonstrated that a mutation in the *AtCPSF73-II* gene was detrimental to the genetic transmission of female gametes. Thus, two 73 kD subunits of the AtCPSF complex appear to have special functions during flower development. The important roles of mRNA 3′-end processing machinery in modulating plant development are discussed.

Keywords mRNA 3′-end processing · Polyadenylation · CPSF complex · Male sterility.

Abbreviations

AD	activation domain
BD	DNA binding domain
CF	cleavage factor
CPSF	cleavage and polyadenylation specificity factor
CstF	cleavage stimulatory factor
CTD	C-terminal domain
DEX	dexamethasone
GFP	green fluorescence protein
GST	glutathione transferase
GUS	β-glucuronidase

Gene accession numbers associated with this paper: AY140902, AY140900, AY168923, AY140901

Electronic Supplementary Material Supplementary material is available for this article at http://dx.doi.org/10.1007/s11103-006-0051-6

R. Xu · H. Zhao · G. Carberry · Q. Q. Li (✉)
Department of Botany, Miami University, Oxford OH 45056, USA
e-mail: liq@muohio.edu

X. Cheng
Department of Microbiology, Miami University, Oxford, OH 45056, USA

R. D. Dinkins
USDA–ARS Forage-Animal Production Research Unit, Lexington KY40546, USA

PAP	poly(A) polymerase
RNAi	RNA interference
RNAP II	RNA polymerase II

Introduction

In eukaryotes, messenger RNA precursors (pre-mRNAs) undergo a series of tightly coupled processing events in the nucleus, including capping at the 5′-end, splicing of introns, and adding a poly(A) tail at the 3′-end. Mature mRNAs are then transported to the cytoplasm where they serve as templates for protein synthesis. Previous studies indicated that polyadenylation plays an important role in mRNA transport from the nucleus into the cytoplasm (Vinciguerra and Stutz 2004), mRNA stability (Gutierrez et al. 1999), transcriptional initiation/elongation and termination (Proudfoot 2004), and translational initiation (Kuersten and Goodwin 2003). This process can also modulate the selection of poly(A) sites, and is a critical point for gene expression regulation during development (Takagaki et al. 1996; Simpson et al. 2003).

The machinery of mRNA 3′-end formation has been extensively studied in mammals and yeast, but little work has been done in plants. The overall process of 3′-end formation is generally similar in mammals and yeast, both of which occur in two tightly coupled reactions in vivo. The pre-mRNA is first subjected to site-specific endonucleolytic cleavage and subsequently a poly(A) tail is synthesized at the 3′-end of the upstream cleavage product. To efficiently accomplish the cleavage and polyadenylation reactions, a large set of *trans*-acting factors, many of which are multi-subunit protein complexes, are required to act in concert to recognize the *cis*-acting signals on the pre-RNA and conduct accurate processing (reviewed by Zhao et al. 1999). In mammals, an efficient in vitro reconstitution of 3′-end formation requires at least six *trans*-acting factors: cleavage and polyadenylation specificity factor (CPSF), cleavage stimulatory factor (CstF), cleavage factor I and II (CF Im and CF IIm), poly(A) polymerase (PAP) and poly(A) binding protein II (PABP2). Of these factors, CPSF, CstF, CF I, CF II and PAP are required for the cleavage reaction, and CPSF, PAP and PABP2 are required for the polyadenylation reaction. However, RNA polymerase II (RNAP II) appears to be an essential 3′-end processing factor in vivo in mammalian cells (Ryan et al. 2002).

While the *cis*-elements of 3′-end formation used by mammals and yeast are rather different, many of the *trans*-acting factors exhibit high homology (Zhao et al. 1999). This coincidence underscores the evolutionally conserved functionality and the important role of mRNA processing. The mammalian CPSF complex consists of the CPSF160, CPSF100, CPSF73, and CPSF30 subunits that are named based on the size (kD) of the polypeptides (Murthy and Manley 1992; Zhao et al. 1999). During transcription, CPSF is first associated with the promoter region by binding to the general transcription factor TFIID. CPSF is then transferred to the C-terminal domain (CTD) of the initiating RNAP II where it traverses the nascent pre-mRNA and joins the polyadenylation complex upon reaching the poly(A) site (Dantonel et al. 1997; Proudfoot 2004). CPSF is not only required for both cleavage and polyadenylation reactions, but also plays an important role in the splicing of terminal introns in vivo (Li et al. 2001). This role is also critical in cytoplasmic polyadenylation (Dickson et al. 1999). Thus, the CPSF complex may be a key player of the transcription and RNA processing factory (Bentley 2002). All four subunits (160, 100, 73 and 30 kD) of the CPSF complex are highly conserved polypeptides, and its yeast homologues (Yhh1/Cft1, Ydh1/Cft2, Ysh1/Brr5, and Yth1) are important components of the CPF complex (Zhao et al. 1999; Ohnacker et al. 2000).

The functions of each of the subunits in the CPSF complex have not yet been fully elucidated. CPSF160 binds to the highly conserved animal polyadenylation signal AAUAAA (Keller et al. 1991). CPSF30 is a nucleic acid binding protein whose sequence contains five zinc finger motifs and a zinc knuckle motif (Barabino et al. 1997). A *Drosophila* homologue of CPSF30 was found to possess endoribonucleolytic activity (Bai and Tolias 1996, 1998). Hence, CPSF30 and its homologues have been proposed to be the nuclease that cleaves pre-mRNA (Zarudnaya et al. 2002). However, this view has been challenged by some evidence showing that CPSF73 may actually be the endonuclease of the cleavage reaction (Ryan et al. 2004; Dominski et al. 2005a). There is significant sequence homology between CPSF100 and CPSF73 proteins (Jenny et al. 1994, 1996), and the antibodies raised against one protein appear to cross-react with the other (Ryan et al. 2004). The functional significance of this finding has not yet been established.

The mechanism of mRNA 3′-end processing in plants remains poorly understood although some recent progress has been made. One of the interesting aspects of plant mRNA 3′-end formation is that the poly(A) signals are significantly different from those in

mammals (reviewed by Li and Hunt 1997; Loke et al. 2005). Previous studies of *trans*-acting factors include poly(A) polymerase (Tarui and Minamikawa 1989; Li et al. 1998) and poly(A) binding proteins (Belostotsky and Meagher 1993). *Arabidopsis* homologues of the mammalian CstF were recently cloned (Yao et al. 2002). Four genes encoding predicted proteins similar to eukaryotic poly(A) polymerases of yeasts and animals were recently identified in the *Arabidopsis thaliana* genome [*AtPAP(I), (II), (III)*, and *(IV)*], and the non-specific poly(A) polymerase activity for these proteins was demonstrated (Addepalli et al. 2004). The cloning of an *Arabidopsis* homologue of the CPSF 100 kD subunit has been reported; the protein was shown to interact with AtPAP(II) in vitro and in yeast two-hybrid assays (Elliott et al. 2003). In addition, we reported mutant characterization of the *AtCPSF73-II* gene, which encodes a homologue of the CPSF 73 kD subunit. Mutation of this gene caused embryo lethality during early stages of seed development and proved detrimental to the genetic transmission of female gametes (Xu et al. 2004). Here we report the identification of *Arabidopsis* homologues of the CPSF complex and demonstrate the critical role of AtCPSF73-I in plant development.

Materials and methods

Plant material and DEX treatment

The *Arabidopsis thaliana* ecotype Columbia (Col) was used as the wild-type strain in this study. The growth conditions of *Arabidopsis* plants were described previously (Xu and Li 2003).

Dexamethasone (DEX, Sigma), a glucocorticoid derivative, was dissolved in ethanol at 30 mM before use and diluted in the working solution (30 μM DEX and 0.01% (w/v) Tween-20; Aoyama and Chua 1997). Immediately after spraying, the plants were typically covered overnight using a transparent plastic dome before returning to regular growing conditions.

Bioinformatic analysis

GenBank database searching and sequence retrieving were conducted at the NCBI website (http://www.ncbi.nlm.nih.gov/). Multiple sequence alignments were performed using the ClustalW program of MacVector 6.5 (Accelry Inc. San Diego, CA). The phylogenetic tree was constructed based on the multiple sequence alignments and by parsimony methods of PAUP* 4.0 software (Sinauer Associates Inc. Sunderland, MA). These included the use of the heuristic search methods with random stepwise addition, tree-bisection-reconnection (TBR) branch swapping and maximum parsimony setting, as well as 100 bootstrap replications. Percentage of sequence similarity and identity was derived from pair-wise sequence comparisons using the GAP program of the University of Wisconsin GCG software package.

Cloning of *Arabidopsis* CPSF genes

Arabidopsis CPSF genes were cloned by RT-PCR, and their ends were confirmed by 5′- and 3′-RACEs using protocols as previously described (Xu et al. 2004). The full-length cDNAs of these genes were deposited in GenBank as *AtCPSF160* (*Arabidopsis* locus ID At5g51660; GenBank Accession# AY140902), *AtCPSF100* (At5g23880; AF283277), *AtCPSF73-I* (At1g61010; AY140900), *AtCPSF73-II* (At2g01730; AY168923) and *AtCPSF30* (At1g30460; AY140901). Cloning of *AtCPSF100* and *AtCPSF73-II* were described previously (Elliott et al. 2003; Xu et al. 2004). *AtCPSF160* was amplified by a long-range PCR reaction using *Takara LA Tag*™ polymerase (TAKARA, Japan) with primers: 5′-ATGAGTTTCGCGGCCTATAAGATGATGCAT and 5′-TCACAAGAAGCTGGTTCCGACAGAGAGATC. *AtCPSF73-I* coding sequence was first isolated using two primer pairs: 5′-AACATGGCTTCTTCTTCTACTTC/5′-TATATCCACCATAAACATGGCTG and 5′-ATTGACTTCCATCAAACAG/5′-AGAAGCTGAGAGAGGGATTGG. The full-length cDNA was subsequently generated by recombinant PCR. *AtCPSF30* was amplified with primer pair 5′-ATGGAGGATGCTGATGGACTT and 5′-CAGAACCCAATTAAAAACCTTAG.

Gene expression profile

For RT-PCR assays, 2 μg of total RNA for each sample was subjected to cDNA synthesis with oligo-(dT) primer using the RETROscript™ kit (Ambion Inc. Austin, Texas) according to the manufacturer's instructions. Subsequently, equal amount (0.5 μl of cDNA synthesis reaction) of the cDNA from each sample was taken as PCR template to amplify gene transcripts. Primers were designed for *AtCPSF160* (5′-TCCAAGAAGCCCTTTGGTAAAAGT and 5′-TTGCACACAACACAAGAACACCTC), *AtCPSF100* (5′-ATGGGTACTTCGGTGCAA and 5′-TCCGATGATTGTAGTCAA), *AtCPSF73-I* (5′-AACATGGCTTCTTCTTCTACTTC

and 5′-CTATATCCACCATAAACATGGCTG), *AtCPSF73-II* (5′-GAGTGATGGTGGATAGAA and 5′-GCGCCTTCCCTCCGCCAGCAACA) and *AtCPSF30* (5′-ATGGAGGATGCTGATGGACTT and 5′-GCATGCCTGTACCGACAATCAGGACCA). In addition, specific primers (5′-TGGACTCTGGTGA TGGTGTG/5′-TTCTGTGGACAATGCCTGGA) for the constitutively expressed actin gene (*ACT8*) were used as an internal control. The following program was used for PCR: 94°C 2 min; 34 cycles of 94°C 30 s, 55°C 30 s, 72°C 1 min.

The microarray data of the expression profile of the *AtCPSF* genes in wild-type plant tissues of different developmental stages were downloaded and extracted from web site http://jsp.weigelworld.org/atgen/expviz.jsp. The microarray experiments have been described by Schmid et al. (2005) and on the web site, in which following statement was found: ''The entire data set was quantile-normalized using gcRMA. Absolute values are linearized gcRMA values. Normalized values are obtained by normalizing absolute values to median for each gene across all samples.'' A detail description of the samples used in Fig. 1C can also be found in the Supplemental Table 1.

Arabidopsis nucleus isolation and nuclear protein extraction

The previously described method (Li et al. 1998) was used with the following modifications. Leaves of three to four-week-old plants were frozen and ground in liquid nitrogen to fine powder with a mortar and a pestle. All subsequent steps were carried out on ice or at 4°C. About 20 g tissue powder was mixed with 100 ml extraction buffer (250 mM Sucrose, 10 mM PIPES-KOH, pH7.0, 10 mM KCl, 10 mM $MgCl_2$, 0.3% (v/v) Triton X-100, 1 mM DTT, and 0.2 mM PMSF) on ice for 30 min, then filtered through two layers of Miracloth and subjected to centrifugation at 2000 × *g* for 10 min. The pellet was resuspended in 15 ml nucleus washing buffer (NWB: 0.4 M hexylene glycol, 10 mM PIPES-KOH, pH7.0, 10 mM $MgCl_2$, 0.3% (v/v)

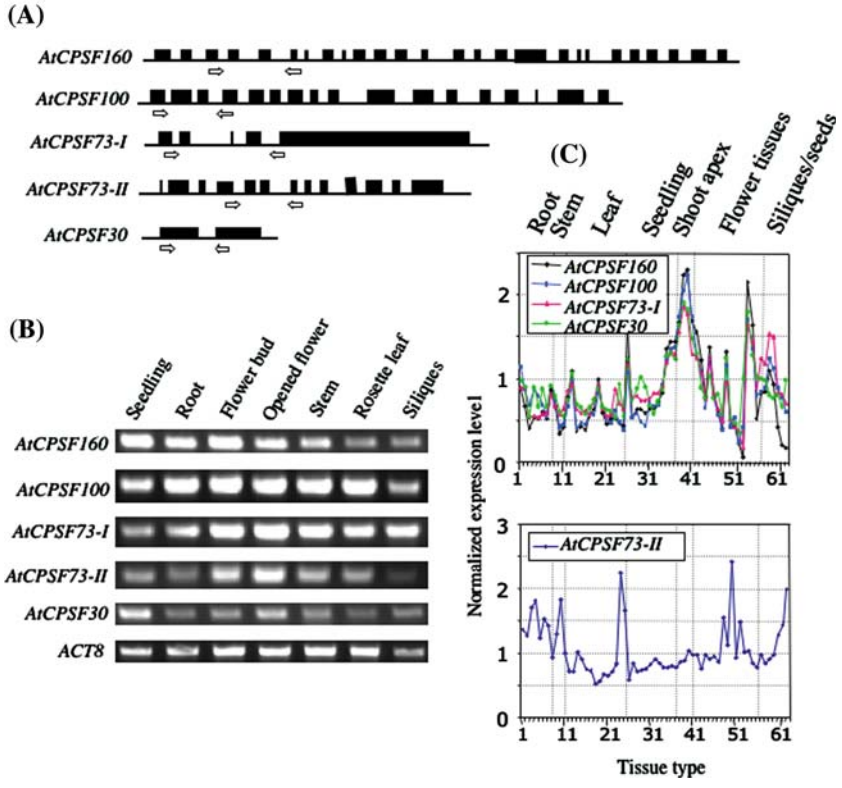

Fig. 1 *Arabidopsis* CPSF genes and their expression profiles. (**A**) Structures of the genes. The dark-filled boxes denote exons, and lines are introns and untranslated regions. The arrows underneath the genes are the primer pairs used for RT-PCR detection in (**B**). (**B**) RT-PCR detection of the expression of the genes in different tissues. The actin8 gene serves as a control. (**C**) Gene expression profiles in different tissues and developmental stages of wild-type plants based on the microarray experiments from the web site (http://jsp.weigelworld.org/atgen/expviz.jsp) as described by Schmid et al. (2005). Each data point represents a different tissue or different stage of development. The organ types, which are divided by the dash lines, are listed on the top. Detail tissue samples used here are listed in the Supplemental Table 1

Triton X-100), filtered through a 100-μm nylon mesh, then layered on 15 ml 30% (v/v) Percoll in NWB and centrifuged again at $1000 \times g$ for 30 min. The pellet was resuspended in 20 ml NWB and filtered by nylon mesh again. Nuclei were further purified in a discontinuous Percoll gradient made by layering 10 ml 30% (v/v) Percoll in NWB on 10 ml floating buffer (0.44 M Sucrose, 80% (v/v) Percoll, 10 mM PIPES-KOH, pH7.0, 10 mM $MgCl_2$). The gradient was centrifuged at $300 \times g$ for 5 min. Most of the nuclei formed a layer just above floating buffer. They were removed, washed once with NWB (without addition of Triton X-100), and finally collected by centrifugation. This nuclei-enriched fraction was resuspended in protein sample buffer (0.175 M Tris–HCl, pH 8.8, 5% (w/v) SDS, 15% (v/v) Glycerol, 30 mM DTT) and subjected to SDS-PAGE and western blot analysis. Protein concentration was determined using the Bradford reagent (Sigma, St. Louis, MO).

Antibodies production and immunoblotting

The antibodies were raised against synthetic peptides from the predicted protein sequences of *Arabidopsis* CPSF genes, i.e. YNHRKERHLNGTVLC for At-CPSF100, CHELEYELNKNSEDN for AtCPSF160, RFFRLYGECREQDC for AtCPSF30, ESE-HSGLKERVRVAFERIQSAV for AtCPSF73-I and SSEAVFLCCNWSIADLELGWEI for AtCPSF73-II. Each peptide, after conjugation with KLH carrier protein using the Imject ® maleimide activated mcKLH kit (Pierce, Rockford, Illinois), was injected into two rabbits (200 μg/injection). Each rabbit was given four booster injections about 2 weeks apart. Peptide-specific antibodies were purified by peptide-BSA conjugant using an affinity-purification procedure as described (Li et al. 1998). Immunoblotting was carried out as previously described (Li et al. 1998). Briefly, the protein gels were transferred to Immobilon-P membranes (Millipore, Bedford, MA) and probed with affinity-purified antibodies. Detection was based on a color reaction of alkaline phosphatase-conjugated goat anti-rabbit antibodies and nitro blue tetrazolium chloride/5-bromo-4-chloro-3-indolyl phosphate. The antibody against Rubisco large subunit was purchased from AgriSera AB (Sweden, www.agrisera.se).

Yeast two-hybrid assay

A Gal4-based two-hybrid system was used as described previously (James et al. 1996). In this system, the yeast strain used was PJ69-4A (*MATa, ade2, trp1-Δ901, leu2-3, 112, his3-200, gal4Δ, gal80Δ, ura3-52, met2::GAL7-lacZ, ADE2:: GAL2-ADE2, LYS2::-GAL1-HIS3*). The expression vectors were pGAD-C(1) and pGBD-C(1) for activation domain (AD) and binding domain (BD), respectively. Gene fusions were constructed by employing the Gateway™ cloning technology (Invitrogen Inc. Carlsbad, CA). First, the entire coding sequences of *Arabidopsis* CPSF genes were mobilized into pDONR 201 vector to generate the Entry Clones. Then the yeast expression vectors, i.e. pGAD-C(1) and pGBD-C(1), were converted into Gateway cloning compatible vectors (provided by Dr. A. G. Hunt), respectively. The cDNA sequences inserted into the Entry Clones were then subcloned into the converted Gateway compatible vectors. DNA sequencing was used to verify the gene fusions. Two-hybrid analyses were carried out essentially as previously described (Gietz et al. 1997). Briefly, the yeast transformants were plated on Synthetic Complete (SC) medium (Clontech, Palo Alto, CA) lacking the appropriate amino acids (Trp, Leu and/or His) for selection purposes. When His selection was used, the Histidine analogue 3-amino-1,2,4-triazole (3-AT) was added to the medium at a concentration of 1 mM to suppress the expression of the reporter gene *HIS3* as recommended (James et al. 1996). Transformants from the empty vectors were used as negative controls, and the yeast genes SNF1 and SNF4 (Fields and Song 1989) were used as a positive control.

Recombinant protein expression, in vitro translation and pull-down assay

The full-length cDNA of *Arabidopsis* CPSF genes and AtPAP(II) as well as GUS gene (a negative control) were subcloned into the Gateway® pDEST20™ vector to express GST fusion proteins using the baculovirus expression system (Invitrogen). GST fusion proteins were purified on glutathione-Sepharose 4B as recommended (Amersham Pharmacia Biotech, Piscataway, NJ). In vitro translations were performed with the TNT® Quick coupled transcription/translation system (Promega, Madison, WI) with the use of RTS wheat germ linear template generation set (Roche, Indianapolis, IN). For GST pull-down, 40 μl GST fusion protein were incubated with in vitro translated proteins overnight at 4°C. The glutathione-Sepharose 4B beads were washed four times with 1 ml 1× PBS for 20 min at 4°C. Proteins were eluted in sample buffer and resolved by SDS-PAGE.

Plasmid construction and plant transformation

Binary vector pCAMBIA1303 (CAMBIA, Australia) was used to make plasmid construct for promoter analysis. A 305 bp DNA fragment of *AtCPSF73-I* native promoter was amplified from *Arabidopsis* genomic DNA by PCR to introduce *Bam*HI and *Nco*I restriction site at the fragment ends and then cloned into pCAMBIA1303 by replacing the CaMV 35S promoter. This native promoter was also fused with the full-length cDNA of *AtCPSF73-I* that was introduced into pPZP211 to generate plasmid pPZP211-P_{CI}-*AtCPSF73-I.* An expression cassette of 35S promoter-*rbcS* 3′end from pKYLX71 was first inserted in the multiple cloning sites of pPZP211, and then the full-length coding sequence of *AtCPSF73-I* was placed in the sense orientation between the 35S promoter and *rbcS* 3′end to generate plasmid pPZP211-P_{35S}-*AtCPSF73-I*. The full-length cDNA of *AtCPSF73-I* was also introduced into a glucocorticoid-mediated transcriptional induction system (Aoyama and Chua 1997) to generate plasmid pTA7001-$P_{GAL4\text{-}UAS}$-*AtCPSF73-I*.

A construct for RNAi gene silencing was made according to previous studies (Wesley et al. 2001). Specifically, a 548 nt coding sequence of *AtCPSF73-I* was amplified by primers 5′-TGACTTCCATCAAACAGTT/5′-ATCGATGCTGTTTAATGGAG, two copies of which were then ligated to both sides of a plant intron sequence in an inverted-repeat orientation. This composite sequence was introduced into pKYLX71 binary vector under control of the CaMV 35S promoter, generating plasmid pKYLX-P_{35S}-*AtCPSF73-I*-RNAi that served a purpose to produce constitutive RNAi interference in transgenic plants. The same composite was also introduced into the glucocorticoid-mediated transcriptional induction system as described above, generating pTA7001-$P_{GAL4\text{-}UAS}$-*AtCPSF73-I*-RNAi which would produce double-stranded RNA for RNAi after induction by DEX in transgenic plants.

All the resulting binary constructs were introduced into *Agrobacterium tumefaciens* and transformed *Arabidopsis thaliana* ecotype Columbia as previously described (Xu and Li 2003).

Transient expression assay and promoter::GUS assay

The full-length cDNA of *AtCPSF73-I* or *AtCPSF73-II* was fused with GFP, resulting in plasmids pKYLX80-P_{35S}-*AtCPSF73-I*-GFP and pKYLX80-P_{35S}-*AtCPSF73-II*-GFP, respectively. The 35S promoter in pKYLX80-P_{35S}-*AtCPSF73-I*-GFP was replaced by the native promoter of *AtCPSF73-I* to generate pKYLX80-P_{CI}-*AtCPSF73-I*-GFP. These plasmids were delivered into tobacco cells and/or onion epidermal cells by microprojectile bombardment and visualized by confocal microscopy as previously described (Dinkins et al. 2003). All experiments were carried out in triplicate and independently performed at least two times. The GUS histochemical staining assay was described previously (Xu and Li 2003).

Results

Arabidopsis CPSF homologues

In the *Arabidopsis* genome, five DNA sequences were identified to encode CPSF homologues. This was accomplished by using the protein sequences of mammalian CPSF subunits and their yeast homologues to query GenBank databases with the BLAST search tools. Three of these sequences encoded homologues to the mammalian CPSF160, 100, and 30 kD subunits. Each of the two distinct sequences encoded a homologue to the 73 kD subunit. Here we designated these genes as *AtCPSF160*, *AtCPSF100*, *AtCPSF73-I*, *AtCPSF73-II*, and *AtCPSF30* following the mammalian nomenclature. The structure of the genes are shown in Fig. 1A. All of these genes were found to express (with some variations in levels) in all tested tissues of *Arabidopsis* by RT-PCR (Fig. 1B) using gene-specific primers spanning gene regions that included one or more introns (Fig. 1A). The RT-PCR results from *Arabidopsis* organs are generally in agreement with the gene expression profiles from a set of microarray experiments using wild-type (Col-0) as described by Schmid et al. (2005). As shown in Fig. 1C, *AtCPSF160, 100, 73-I*, and *30* are expressed higher in seedling, shoot apex, carpel, and silique with young seeds, respectively. Interestingly, all these four genes have a lower expression in stamen or mature pollen (data point 51 or 52 in Fig. 1C). However, the expression of *AtCPSF73-II* is outstanding from all others in which it has distinct peaks that are different from the rest. In particular, it has high expression level on senescence leaf, petal, stamen, pollen and late stages of siliques with seeds (Fig. 1C, bottom graph). The fact that *AtCPSF73-II* has high expression level in floral tissues (including siliques) is in good agreement with our previous northern blot results (Xu et al. 2004). The expression of *AtCPSF73-I* and *II* are in direct contrast when senescing leaf, sepal, mature pollen / stamen and late stage silique are concerned: *AtCPSF73-I* down, and *AtCPSF730-II* up. These results

imply that *AtCPSF73-I* and *II*, although with conserved sequences (see below), each may have special functions at different stages of development.

GenBank database searches identified proteins that are most similar to these *Arabidopsis* CPSF homologues from various organisms or species. As expected, the resulted homologues included all of the CPSF-related proteins from mammals and yeast that have been characterized for their biochemical function in the polyadenylation process. Since the mammalian CPSF73 and CPSF100 share significant homology at the protein level (Jenny et al. 1994, 1996), a phylogenetic tree was constructed from all of the AtCPSF100, AtCPSF73-I, and AtCPSF73-II homologues (Fig. 2). These homologues were clearly clustered into three distinct groups as represented by AtCPSF100, AtCPSF73-I, and AtCPSF73-II. AtCPSF73-II clustered with a group of functionally unknown proteins from diverse organisms including mammals, insects, fungi, a nematode, and a *Plasmodium*. No homologue was found against GenBank's yeast sequences, including *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. It is noteworthy that the AtCPSF73-I and AtCPSF73-II groups collectively formed

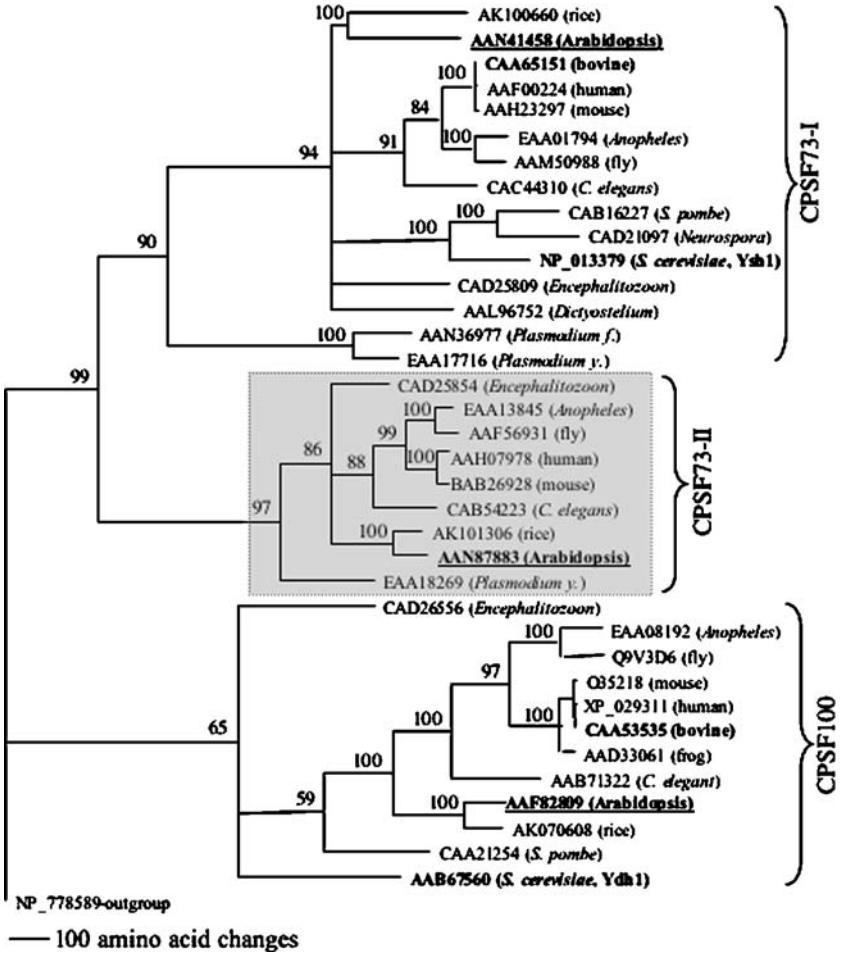

Fig. 2 A phylogenetic tree of AtCPSF100, AtCPSF 73-I, AtCPSF73-II and their closest homologues from various organisms. These homologues, denoted by GenBank accession number with its organism source in parentheses, fall into three distinct groups. AtCPSF100 represents a group of the 100 kD subunit of CPSF homologues. AtCPSF73-I represents a group of the 73 kD subunit of CPSF homologues identified so far. AtCPSF73-II represents a group (which is highlighted by the shaded box) of novel proteins, with unknown function, from diverse organisms, but is closely related to the 73 kD subunit of CPSF. The parsimonious trees was generated by PAUP* software. A scale bar is placed at the bottom of tree for the horizontal branch lengths reflecting the number of amino acid changes between taxa. The numbers (bootstrap values) adjacent to nodes are the percentages of 100 heuristic bootstrap trials in which the indicated protein groups were found. *Arabidopsis* CPSF proteins are highlighted by bold font and underlined. In bold are those proteins previously characterized to be involved in the CPSF complex. The full names of the species involved are as following: human (*Homo sapiens*), bovine (*Bos taurus*), rice (*Oryza sativar*), fly (*Drosophila melanogaster*), frog (*Xenopus laevis*), mosquito (*Anopheles gambiae*), nematode (*Caenorhabditis elegans*), fungi (*Encephalitozoon cuniculi, Neurospora crassa*), yeast (*Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*), Amoeba *(Dictyostelium discoideum)*, *Plasmodium falciparum*, and *Plasmodium yoelii*

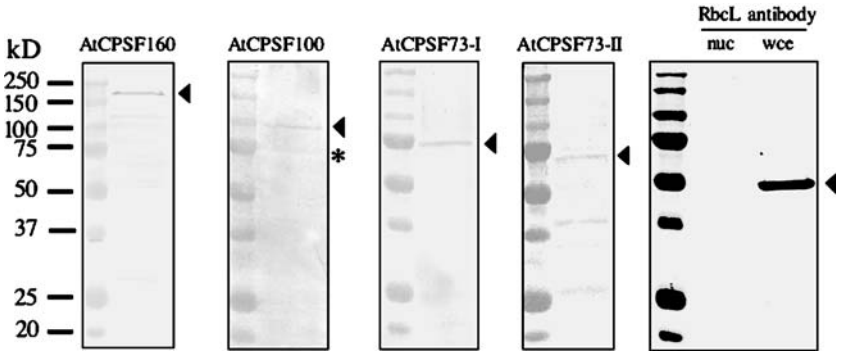

Fig. 3 Immunodetection of *Arabidopsis* CPSF proteins. *Arabidopsis* nuclear extracts (~20 μg protein/lane) were probed by the affinity-purified peptide-specific antibodies against AtCPSF160, AtCPSF100, AtCPSF73-I and AtCPSF73-II as indicated above each panel. The arrowheads indicate the expected bands. The asterisk "*" indicates a cross-reaction band corresponding to AtCPSF73-I. Protein size markers are on the left lane of each panel. For Rubisco large subunit (RbcL) antibody panel (as a control), equal amount of protein (10 μg) was loaded to nuclear (nuc) and whole cell extract (wce) lanes, respectively

a higher-order cluster distinct from the AtCPSF100 group, therefore suggesting a closer evolutionary relationship between AtCPSF73-I and AtCPSF73-II homologues. Further analysis of sequences indicated an average of 33.8% identity (43.0% similarity) between proteins of the AtCPSF73-I group and the AtCPSF73-II group. An average of 57.4% identity (67.4% similarity) within the AtCPSF73-I group and 59.3% identity (69.0% similarity) within the AtCPSF73-II group was observed. This result supports a distinct relationship between the AtCPSF73-I group and the AtCPSF73-II group. Taken together, we suggest that AtCPSF73-I represents the counterpart of mammalian CPSF73 and that AtCPSF73-II is a novel protein evolutionarily related to CPSF73.

Antibodies against synthetic peptides of these homologues were produced to probe *Arabidopsis* nuclear extracts by western blots. As illustrated in Fig. 3, the peptide-affinity purified specific antibodies against AtCPSF160, AtCPSF73-I, and AtCPSF73-II recognized a major band of expected protein sizes of 158 kD, 77 kD, and 68 kD, respectively. The anti-AtCPSF100 antibodies detected a major band somewhat larger than the expected band for AtCPSF100 (82 kD), but cross-reacted slightly with a band corresponding to AtCPSF73-I. Such a cross immunorecognition of CPSF73 and CPSF100 was also noted in the mammalian extract (Ryan et al. 2004). The peptide from AtCPSF30 generated antibodies that cross-reacted with a number of proteins (data not shown); however, a 28 kD band, the predicted size of *AtCPSF30*, was absent in a T-DNA insertion mutant (*oxt6*) (Delaney et al. 2006). To ensure the quality of the nuclear extract (free of cytoplasmic contamination), a control western blot was performed with an antibody against Rubisco large subunit that should only be found in the cytoplasmic or whole cell extract, not in the nuclear extract. As shown in Fig. 3, this antibody detected no band in the nuclear extract. From these studies, it was noted that the AtCPSFs are all low-abundance nuclear localized proteins in *Arabidopsis*.

Interactions among *Arabidopsis* CPSF subunits and with PAP

The ability of the *Arabidopsis* CPSF homologues to interact with each other to form a protein complex, a characteristic of the CPSF complex in both bovine and yeast (Zhao et al. 1999), was characterized using a yeast two-hybrid system. When these genes were fused with the DNA binding domain (BD) or the activation domain (AD) and transformed into yeast, most of the resulting transformants did not grow on the selection medium (Fig. 4A). This indicated that no self-activation occurred. However, the transformants carrying the BD-AtCPSF30 fusion grew normally, while the one carrying the AD-AtCPSF100 fusion grew weakly (Fig. 4A). These results suggest the self-activation activities of the gene fusions. This was taken into consideration when determining protein–protein interactions by pair-wise combinations in the AD and BD gene fusions described below. In some experiments, a plasmid containing a truncated form of the AtCPSF100 was fused in the AD vector; this plasmid alone had no detectable binding activity (Elliot et al. 2003). In addition, AtPAP(II) and AtPAP(III) (Hunt et al. 2000; Addepalli et al. 2004) were also investigated to determine their associations with the *Arabidopsis* CPSF complex.

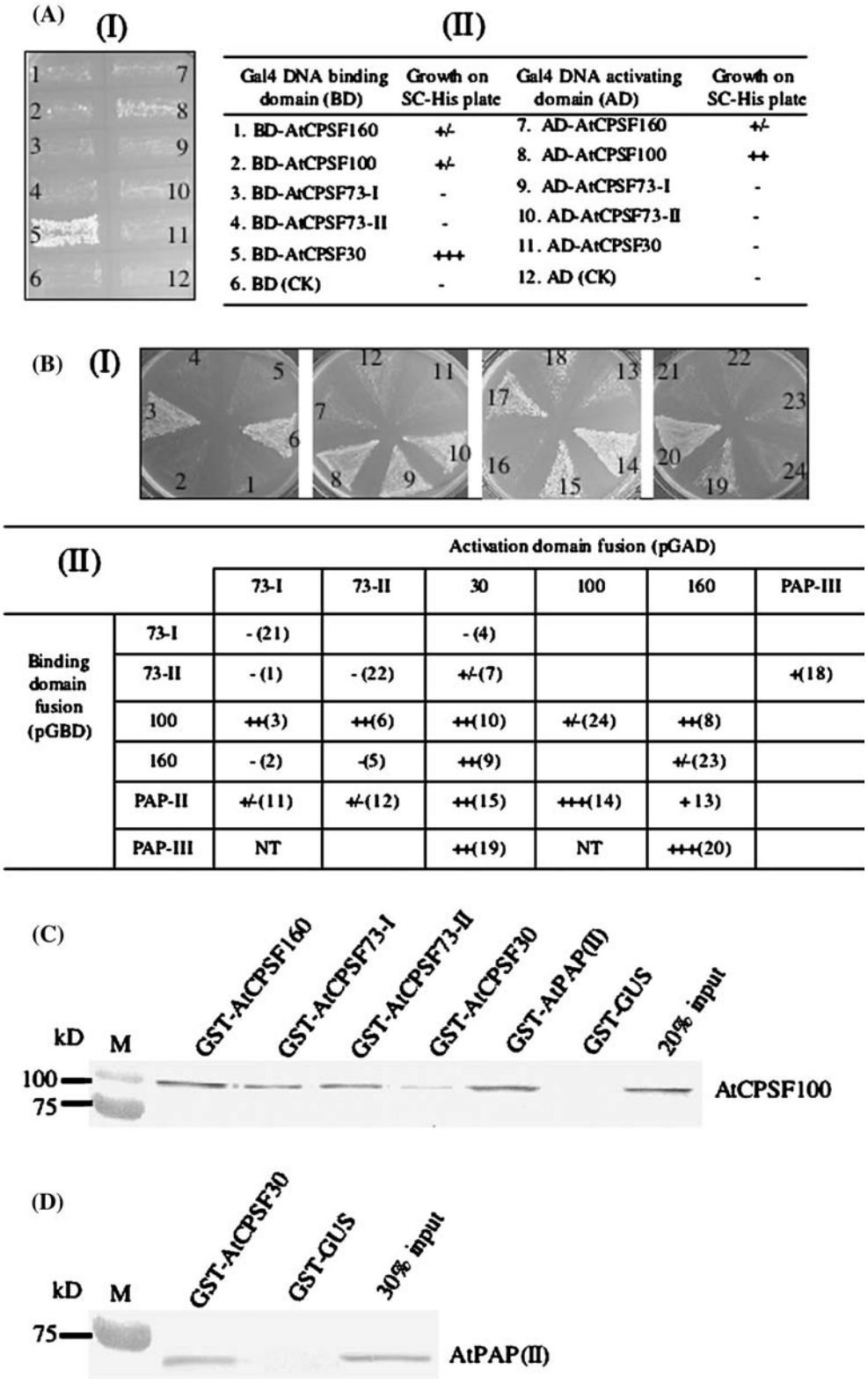

Fig. 4 Protein–protein interactions of *Arabidopsis* CPSF proteins as well as PAP. (**A**) and (**B**), yeast two-hybrid assays. The numbers in the picture (I) correspond to the numbers in (II). (A) self-activation tests of the activation domain (AD) and binding domain (BD) fusion with AtCPSF. The gene fusion was transformed into yeast strain PJ69-4A and then grown on SC-His selection medium. (**B**) Pair-wise protein–protein interactions. The AD fusion and BD fusion were co-transformed into yeast and then grown on SC-Trp-Leu-His selection medium. The relative growth of transformed yeast cells on the selection medium (as seen in the pictures) were scored in the corresponding tables as follow: +++, strong growth; ++, good growth; +, poor growth; +/−, very little growth; −, no growth; NT, not tested. (**C**) and (**D**) Protein pull-down assays verifying interaction results in (**B**). Experiments were performed with the GST fusion proteins and in vitro translated native proteins as indicated to the right of each panel. The fusion of GST-GUS serves as the control. The lane labeled with ''20% or 30% Input'' shows 20% or 30% of in vitro translated protein in the assay. AtPAP(II) in vitro translation produced a truncated proteins of ~70 kD in (**D**)

Pair-wise interactions are depicted in Fig. 4B. AtCPSF100 appears to interact with all of the proteins tested. Both AtCPSF73-I and AtCPSF73-II showed a similar interaction pattern with AtCPSF100, and potentially with AtPAPs. AtCPSF160 and AtCPSF30 interacted with each other. No interaction was

detected for the following combinations: AD-AtCPSF73-I/BD-AtCPSF73-I, AD-AtCPSF73-II/BD-AtCPSF73-II, and AD-AtCPSF160/BD-AtCPSF160. These results suggest that AtCPSF73-I, AtCPSF73-II, and AtCPSF160 proteins may not form homodimers.

We further confirmed these interactions via direct protein–protein interactions by pull-down assays with in vitro expressed proteins and GST fusion proteins produced using a baculoviral-insect cell expression system. The protein pairs were incubated, and then pulled-down with glutathione Sepharose 4B beads. After washing, the bound proteins were denatured and resolved by SDS-PAGE; the blots were detected using antibodies against AtCPSF 100 and AtPAP(II) (a gift from A.G. Hunt). As shown in Fig. 4C, AtCPSF100 was pulled-down by the fusion proteins of GST with AtCPSF160, AtCPSF73-I, AtCPSF73-II, AtCPSF30, and AtPAP(II), respectively. The interaction between GST-AtCPSF30 and AtPAP(II) is demonstrated in Fig. 4D.

Both AtCPSF73-I and -II are localized in the nucleus

The detection of the AtCPSF73-I and II proteins by western blot analysis indicated that they were present in the nuclear extract (Fig. 3). To confirm this finding, we fused both AtCPSF73-I and II to Green Fluorescent Protein (GFP) in order to analyze its localization using transient expression assays in both onion and tobacco cells. GFP fluorescence from transformed cells of the AtCPSF73-II-GFP fusion was observed in the nuclei of both onion epidermal cells and tobacco leaf cells (Fig. 5A–F), indicating a nuclear localization of AtCPSF73-II in plants. The AtCPSF73-I-GFP fusion protein was more difficult to visualize due to the low level fluorescence observed. It appeared to cause a lethal phenotype, as described below. Figure 5 (panels G through I) shows one of such localization experiments leading to the conclusion that AtCPSF73-I is also nuclear localized. In order to see the weak GFP

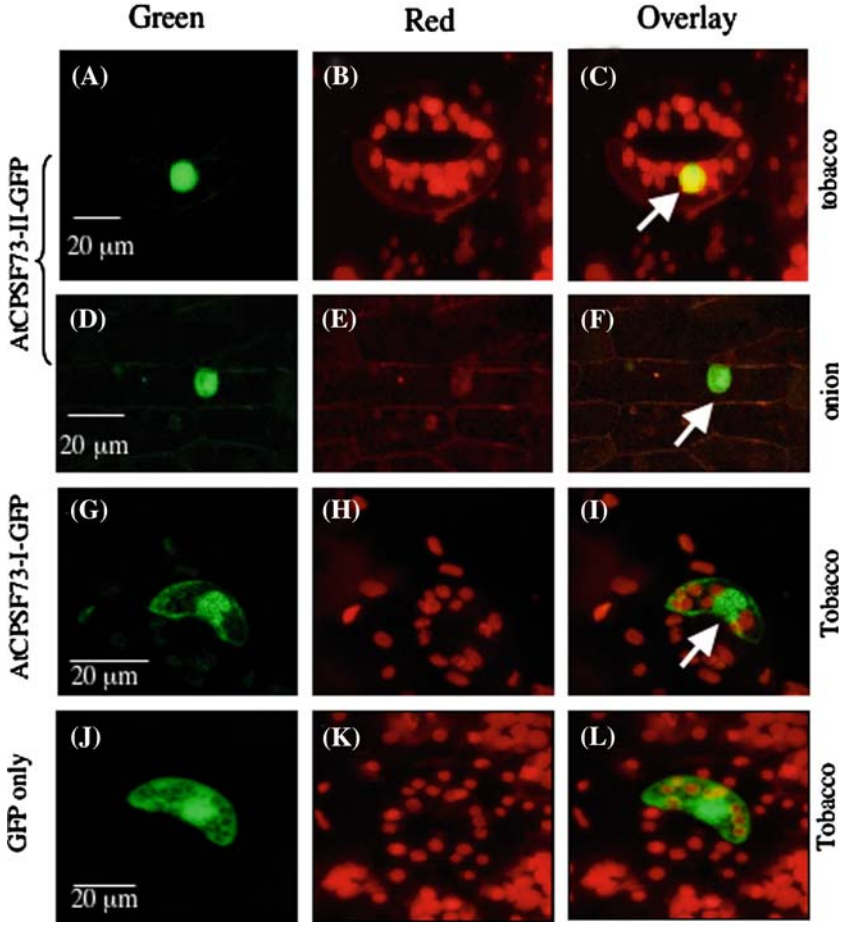

Fig. 5 Nuclear localization of AtCPSF73-I and -II. The fusions of cDNA with the reporter GFP were delivered into both tobacco leaf and onion epidermal cells (*AtCPSF73-II*, **A** to **F**), or just tobacco leaf cells (*AtCPSF73-I*, **G** to **I**; or GFP only, **J** to **L**), respectively, by particle bombardment. As shown in the overlay of images acquired at the same confocal plane using Texas red (red) or FITC (green) filters, GFP fluorescence is located in the nucleus (indicated by arrows). The non-fusion GFP (**J** to **L**) is distributed both in the nuclear and the cytoplasm

fluorescence, the photomultiplier setting was turned high thus autofluorescence of other cellular components (e.g. cellular membrane) were observed (compare the background of Fig. 5G and J). However, the majority of the fusion protein was retained in the nucleus (Fig. 5I), as compared to the non-fusion GFP control that was evenly distributed throughout the cytoplasm and nucleus (Fig. 5J–L). Similar results were obtained when the AtCPSF73-I-GFP fusion was driven by the native promoter of AtCPSF73-I (the construct is described below).

Both knockdown and over-expression of *AtCPSF73-I* in plants cause lethality

We investigated the effects of *AtCPSF73-I* gene silencing in *Arabidopsis* using RNA interference (RNAi) methodology (Wesley et al. 2001). When an inverse repeat sequence of a part of *AtCPSF73-I* gene (leading to RNAi) was expressed under the control of the CaMV 35S promoter (pKYLX71-P_{35S}-*AtCPSF73-I*-RNAi), no transgenic plants were obtained after screening about 30,000 seeds, suggesting that this construct resulted in lethality. In contrast, when the RNAi construct was expressed under the control of the glucocorticoid inducible promoter system (Aoyama and Chua 1997; pTA7001-$P_{GAL4\text{-}UAS}$-*AtCPSF73-I*-RNAi), more than 50 transgenic *Arabidopsis* plants (T_0) were obtained. When the transgenic plants (homozygous lines) were sprayed with 30 μM dexamethasone (DEX) to induce RNAi, almost no *AtCPSF73-I* expression was detected by northern blot (Fig. 6C), suggesting an efficient knockdown of this gene. Subsequently, the effects of the *AtCPSF73-I* knockdown at different stages of plant development were examined. The DEX imbibing experiment indicated that DEX treatment inhibited seed germination from these transgenic plants (data not shown). DEX treatment, applied once at any stage of plant development, caused the plants to gradually turn yellow and eventually die (representative examples are shown in Fig. 6A & B), whereas the vector-transformed control plants remained normal.

The change at the AtCPSF73-I protein level upon RNAi was also monitored by western blot. Figure 6D shows that the AtCPSF73-I protein was significantly reduced when plants visibly yellowed one day after DEX treatment. It was noted that the AtCPSF73-I protein level fluctuated somewhat before turning lower, possibly due to a mechanism of protein expression self-regulation.

Similarly, the effect of AtCPSF73-I over-expression in plants was also evaluated. No transgenic plants were recovered that contained the construct pPZP211-P_{35S}-*AtCPSF73-I*, in which the full-length cDNA of *AtCPSF73-I* was under the control of the CaMV 35S promoter. In contrast, more than 60 transgenic plants were successfully generated with the pTA7001-$P_{GAL4\text{-}UAS}$-*AtCPSF73-I* construct. Note that this differs from the RNAi construct described above, in which the full-length cDNA is under the control of the glucocorticoid inducible promoter. When subjected to DEX spraying for four consecutive days (30 μM, once per day), the homozygous line transgenic plants turned yellow and eventually died (data not shown). Western blot analysis indicated that the steady-state level of the AtCPSF73-I protein concentration upon DEX induction gradually increased until the fourth day when the plants turned yellow and then subsequently declined (Fig. 6E).

These results collectively indicate both the importance of *AtCPSF73-I* in *Arabidopsis*, and the criticality of the control of expression levels of *AtCPSF73-I* for *Arabidopsis* growth and development.

Native promoter driven *AtCPSF73-I* cDNA expression causes male sterility

Since the overexpression of the *AtCPSF73*-I gene by the CaMV 35S promoter resulted in a lethal phenotype, we expressed the *AtCPSF73*-I gene under the control of its native promoter. Because there are only 305 bp from the stop codon of the previous gene to the AtCPSF73-I ATG start codon, this sequence was cloned and fused with the GUS (*β*-glucuronidase) reporter gene and then transformed into *Arabidopsis*. GUS staining of the transgenic plants indicated that this native promoter governed GUS expression throughout the whole plant (Fig. 7A).

When the 305 bp promoter sequence was fused with the full-length cDNA of *AtCPSF73-I* (construct pPZP211-P_{CI}-*AtCPSF73-I*) and then transformed into *Arabidopsis*, transformants were successfully obtained. However, the primary transgenic plants were male sterile. Fifty-one of the 65 (78.4%) transgenic plants produced no siliques (Fig. 7B), while the remaining 14 (21.6%) produced only a few short siliques on each plant. The AtCPSF73-I protein level in these transgenic plants was slight elevated (about 22% more than control; Fig. 7F), indicative that the transgene did contribute to the expression level.

Microscopic examination revealed that the anthers were unable to dehisce even though the flowers had

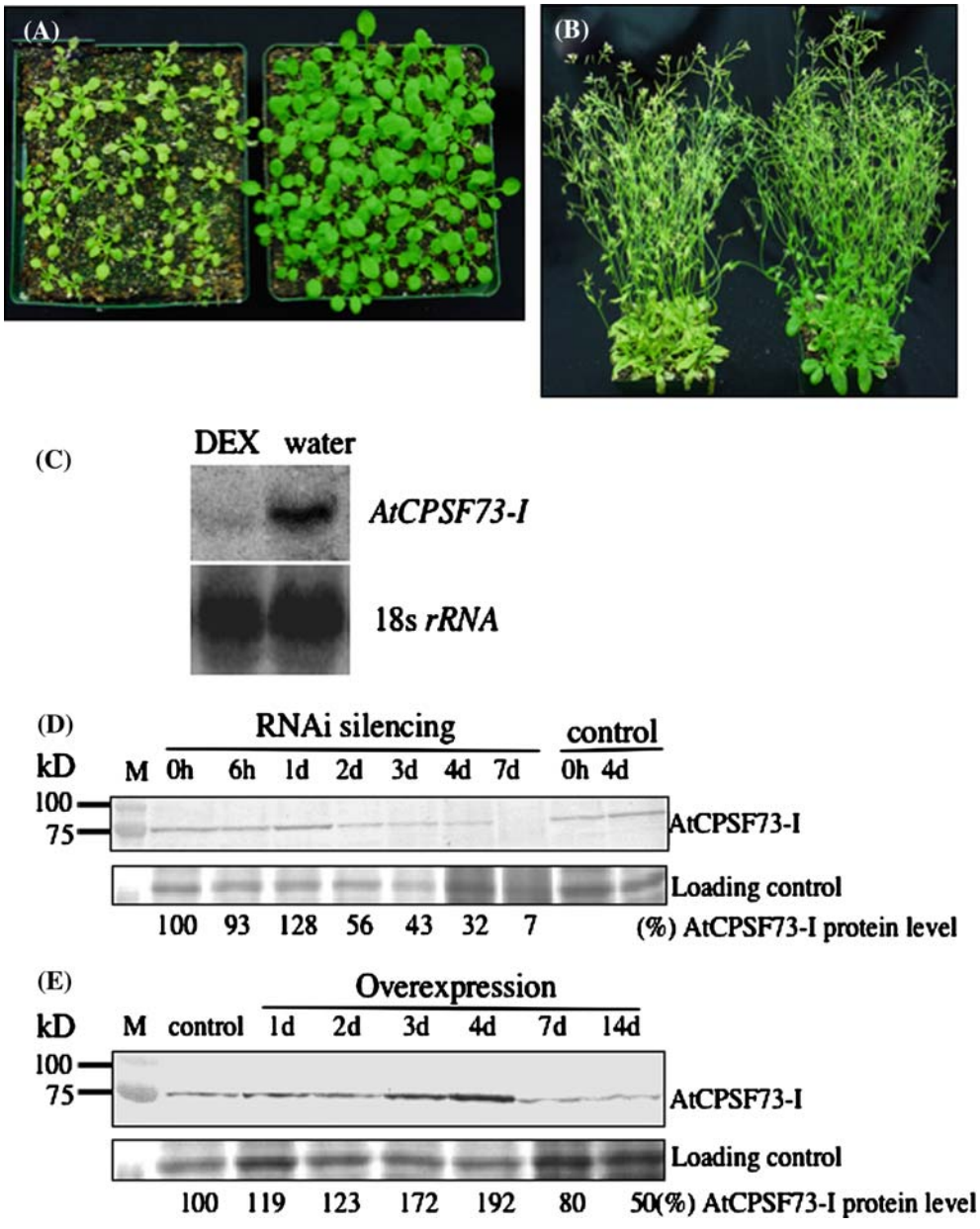

Fig. 6 The lethal effects of both DEX-induced RNAi gene silencing and over-expression of AtCPSF73-I in plants. (**A**) and (**B**) Showing the lethal effect of DEX-induced RNAi gene silencing of AtCPSF73-I in plants. The transgenic plants carrying plasmid pTA7001-$P_{GAL4\text{-}UAS}$-AtCPSF73-I-RNAi (the pots on the left) turned yellow one day later after 30 μm DEX spray and eventually died, compare to vector transformed control plants (the right pots) which were not impacted by the spray. The age of plants was 19 days in (**A**) and 40 days in (**B**). (**C**) A northern blot (~50 μg of total RNA/lane) probed by *AtCPSF73-I*, showing greatly reduced *AtCPSF73-I* mRNA in the transgenic line treated by 30 μm DEX to induce RNAi, or by spraying water as a control. The same blot was tested by a 18s rRNA probe to show equal loading. (**D**) Effect of DEX-induced RNAi silencing on AtCPSF73-I protein levels. Transgenic plants for inducing RNAi of *AtCPSF73-I* were sprayed with 30 μM DEX, then sampled at 0 (just before treatment), 6 h, and 1, 2, 3, 4, 7 d after treatment for extracting nuclear proteins (20 μg/lane) which was probed by anti-AtCPSF73-I antibodies. The relative percentage (%) of AtCPSF73-I protein corrected for loading (see a lane of coomassie-stained gel below) compared with lane 1 is indicated below each lane. (**E**) Effect of DEX-induced overexpression on AtCPSF73-I protein levels. Similar assays as in (**D**), but the transgenic plants were from the construct pTA7001-$P_{GAL4\text{-}UAS}$-AtCPSF73-I, and subjected to a consecutive DEX spraying for four days (30 μM once per day) when the transgenic plants turned yellow. ''M'' indicates molecular weight markers. Quantification of protein bands was done by ImageQuant™ software (Molecular Dynamics, Sunnyvale, CA)

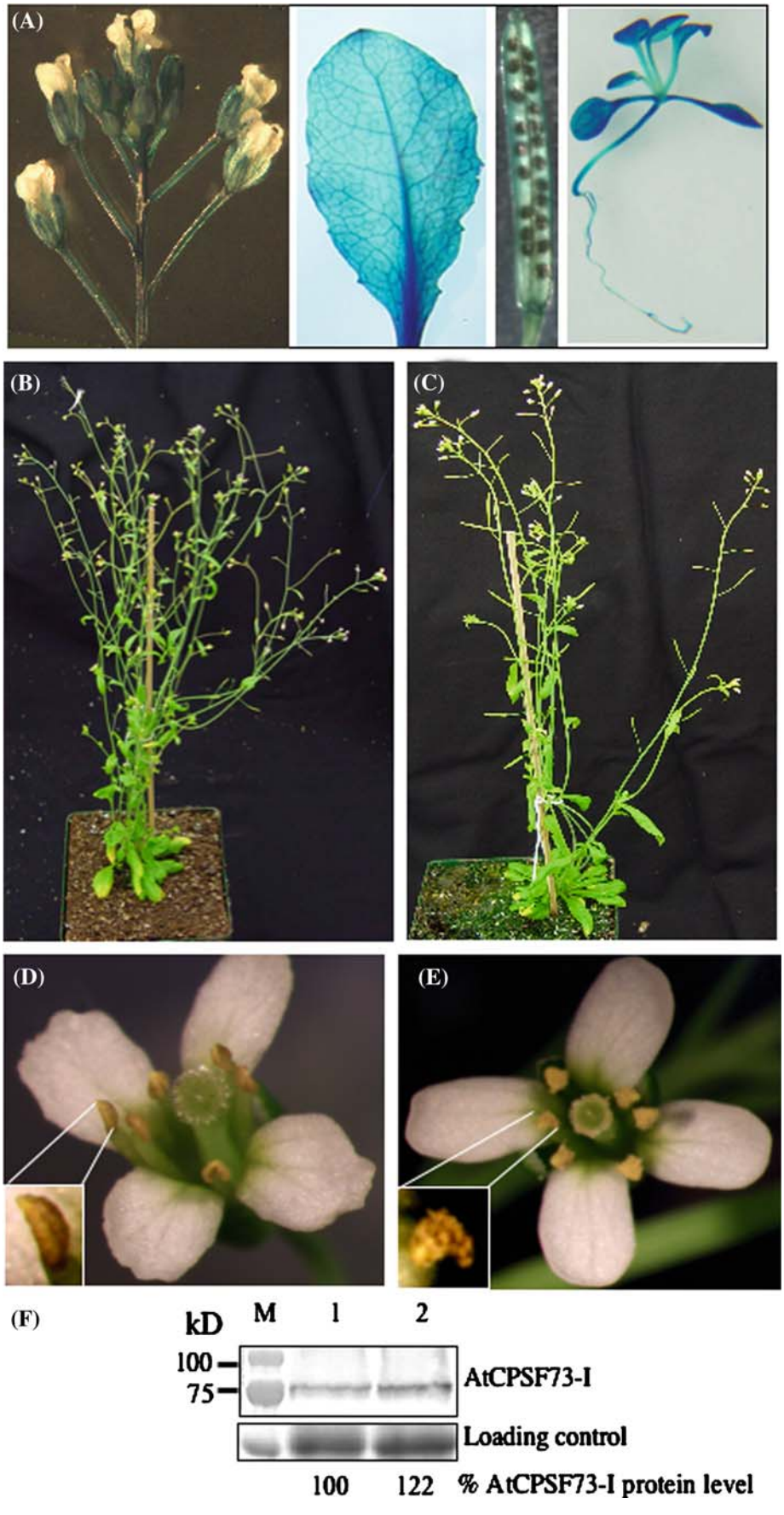

Fig. 7 Impact of the native promoter driven expression of *AtCPSF73-I* gene on plant development. (**A**) GUS activity driven by the *AtCPSF73-I* native promoter. From left to right: flowers, leaf, silique, and seedling. (**B**–**E**) Phenotype analysis of transgenic plants harboring the fusion of *AtCPSF73-I* full-length cDNA under the control of its native promoter and the wild-type control. (**B**) A representative transgenic plant showing sterility. No mature siliques were found, comparing to a normal wild-type plant in (**C**). (**D**) A representative open flower from sterile transgenic plants showing its unopened anther (insert), compared to a wild type flower with open anthers in (**E**). (**F**) A western blot showing AtCPSF73-I protein level in the nuclear extract of a pooled seedling sample of transgenic plant lines. Thirty μg of protein was loaded to each lane. Lane 1, control; lane 2, transgenic plants of construct pPZP211-P_{CI}-*AtCPSF73-I*. The numbers below the ''loading control'' indicate the relative levels of the AtCPSF73-I protein. Another blot with different lines yielded similar result (123% on the transgene over control)

opened (see Fig. 7D, as compared to wild type in Fig. 7E). However, it was noted that these anthers did eventually dehisce a few days later. Pollen viability staining indicated that the pollen from these anthers appeared to be viable (data not shown). When the pollen was collected and applied to the stigmas of the flowers of wild type plants, normal seed set was observed (data not shown), confirming the viability of the pollen. In reciprocal crosses, pollination of the transgenic plants by wild-type pollen also resulted in normal silique development, suggesting the female organ was fertile on the transgenic plants. Male sterility appeared again in the offspring of the cross experiments (data not shown), suggesting that the trait was inheritable. Thus, we conclude that the sterility of the transgenic plants was solely due to the delayed dehiscence of anthers.

While the actual mechanism preventing pollen release in these transgenic lines will require additional analysis, the abnormality in flower development was apparently caused by the expression of at least one extra copy of *AtCPSF73-I* in addition to the native one. This result coincides with our studies of RNAi gene silencing and overexpression as described above, which demonstrate that the control of AtCPSF73-I level is critical, and any disturbance in the protein level results in developmental abnormalities.

Discussion

Messenger RNA 3′-end formation is a fundamental process in eukaryotic transcription and translation, and the CPSF complex plays a critical role during 3′-end processing. A total of five genes, as described in this report, were found to encode the CPSF homologues in the *Arabidopsis* genome. These genes are all expressed in *Arabidopsis* plants and the encoded proteins were all detectable in nuclear extracts. Since protein–protein interactions are important for the formation of protein complexes which carry out biological functions, we established that these *Arabidopsis* CPSF homologues were able to interact with each other using yeast two-hybrid and pull-down assays (Fig. 4). Moreover, we also examined the genetic contributions of the two AtCPSF73 subunits and found that the disturbance of the expression level of the genes can cause profound impact on plant growth and development (Fig. 6). Surprisingly, the presence of an additional copy of the *AtCPSF73-I* gene led to an impact on male sterility by preventing anther dehiscence (Fig. 7).

The interaction profile of AtCPSF subunits is similar to that found among subunits of mammalian CPSF complex or their yeast counterparts (Zhao et al. 1999; Barabino et al. 2000; Kyburz et al. 2003; Dominski et al. 2005b). In mammals, the CPSF complex binds to AAUAAA through CPSF160 with the help of CPSF30 and CPSF100 (Zhao et al. 1999), whereas CPSF73 appeared to interact with CPSF100 (Calzado et al. 2004; Dominski et al. 2005b). A homologue of CPSF100 in yeast, Ydh1/Cft2, was demonstrated to interact with both Yhh1/Cft1 and Ysh1/Brr5 (which are homologous to CPSF160 and CPSF73, respectively). Interaction of Ysh1/Brr5 with Yth1, a yeast homologue of CPSF30, was also observed (Barabino et al. 2000; Kyburz, et al. 2003). Also in mammals, interaction of CPSF, via the 160 kD subunit, with PAP is a critical step for cleavage and specific polyadenylation (Zhao et al. 1999). The *Arabidopsis* genome has four different genes encoding PAP which all show nonspecific poly(A) polymerase activities (Hunt et al. 2000; Addepalli et al. 2004). Interaction of AtCPSF100 with AtPAP(II) has been demonstrated in a previous report (Elliott et al. 2003), and was confirmed in the present study (Fig. 4). Interestingly, we also observed that AtPAP(III) interacted with both AtCPSF160 and AtCPSF30 in the yeast two-hybrid assays. Moreover, the microarray data showed that the four genes of *Arabidopsis* CPSF (160, 100, 73-I and 30) are expressed in more or less synchronized fashion (Fig. 1C). All these data suggest that the *Arabidopsis* CPSF homologues are a part of the polyadenylation apparatus in plants.

There are two closely related CPSF73 homologues, AtCPSF73-I and AtCPSF73-II, encoded by two distinct genes in the *Arabidopsis* genome. The expression of both *AtCPSF73-I* and *AtCPSF73-II* was detected in all *Arabidopsis* tissues by RT-PCR (Fig. 1). However, our northern blot results indicated that *AtCPSF73-II* is expressed mostly in flowering tissues (Xu et al. 2004), while AtCPSF73-I is ubiquitously expressed (Fig. 1). Interestingly, a new set of microarray experiments with a detailed analysis of *Arabidopsis* gene expression profiles of different tissues and developmental stages revealed two distinct expression patterns of these two genes (Schmid et al 2005; Fig. 1C). The following observations support the notion that *AtCPSF73-I* and *-II* are not functionally redundant: a transgene of *AtCPSF73-I* driven by the *AtCPSF73-II* promoter was unable to rescue an *AtCPSF73-II* mutant (Xu et al. 2004); and there was wild-type *AtCPSF73-I* gene in the *AtCPSF73-II* mutant that still showed the embryo arrest phenotype (Xu et al. 2004). Mutation analysis confirmed that both *AtCPSF73-I* and *AtCPSF73-II* are essential in plant development (no homozygous mutants found for *AtCPSF73-I*, Li Q.Q. unpublished

observation; Xu et al. 2004). However, AtCPSF73-I and AtCPSF73-II are highly similar proteins and show very close evolutionary relationship (Fig. 2). AtCPSF73-I seems to represent the counterpart of CPSF73 so far identified in mammals; AtCPSF73-II represents a group of function-unknown homologous proteins present in diverse organisms (Fig. 2). Thus, we reasonably suggest that AtCPSF73-II is a novel CPSF73-related protein.

Both AtCPSF73-I and AtCPSF73-II are localized in the nucleus (Fig. 5), and appear to interact with AtCPSF100 (Fig. 4). This raises a question regarding the relationship between the two AtCPSF73 subunits in the AtCPSF complex in vivo. Presently it is unknown whether both AtCPSF73-I and AtCPSF73-II could reside in the same CPSF complex via binding different sites on AtCPSF100, or if they could form different complexes with AtCPSF100 competing for the same binding site. Recently, a human homologue (AAH07978 in Fig. 2) of AtCPSF73-II, named RC68, was demonstrated to interact with RC74 (which is homologous to CPSF100) through its C-terminal fragment (Dominski et al. 2005). The homology between AtCPSF73-I and II is significant throughout the protein, except at the C-terminal portion. Thus, if interaction with CFSF100 is in the C-terminal region, it may be that the two proteins interact differently with CPSF100.

The biochemical function of the 73 kD subunit of CPSF got some attention recently. There is some evidence that suggests that the CPSF73 may be the actual endonuclease for the cleavage reaction during mRNA 3′-end formation (Ryan et al. 2004). Moreover, Dominski et al. (2005a) have recently shown that CPSF73 appear also to be the endonuclease for histon-pre-RNA processing. Our present studies demonstrated that AtCPSF73-I is essential in plant development since RNAi-mediated knockdown of AtCPSF73-I caused plant lethality (Fig. 6). In addition, over-accumulation of AtCPSF73-I under the control of the CaMV 35S promoter also appears to be lethal (Fig. 6). This suggests that the control of AtCPSF73-I expression level is required for normal growth and development of plants. It was documented that the control of poly(A) polymerase level is essential to cytoplasmic polyadenylation and early development in *Drosophila* (Juge et al. 2002). There is increasing evidence underlying the important role of expression levels of polyadenylation factors in development. For example, during B-cell differentiation, the level of CstF64 was increased to regulate alternative processing of IgM heavy chain pre-mRNA (Takagaki et al. 1996). The mRNAs of CPSF160 and CstF64 were greatly overexpressed in mouse testicular RNA than in liver RNA (Dass et al. 2001). The mechanism of such a control may play an important role in modulating development. This argument is supported by a recent report which showed that the human immunodeficiency virus type 1 (HIV-1) Tat protein modulated CPSF73 for regulating gene expression. It was found that HIV-1 Tat protein specifically increased the expression of CPSF73 of the host cell, and CPSF73 exerts transcriptional activities by repressing the *mdm2* gene promoter (Calzado et al. 2004). In fact, the entire process of mRNA 3′-end formation is finely coordinated by a large number of proteins through which an extensive protein–protein interaction network is formed, coupling with other fundamental processes such as transcription, capping, splicing, and translation (Bentley 2002; Proudfoot 2004).

Examples demonstrating the significant role of mRNA 3′-end formation in plant developmental modulation have been described. In *Arabidopsis*, the mRNA 3′-end processing factor FY, related to Pfs2p of *S. cerevisiae*, was found to interact with FCA which is a nuclear RNA binding protein controlling flowering time (Simpson et al. 2003). The FCA/FY interaction is not only required for regulating FCA expression through pre-mRNA processing, but also for the down-regulation of the floral repressor *FLC* (Simpson et al. 2003). Two RNA binding proteins, HEN4 and HUA1, also act in floral morphogenesis by specifically promoting the processing of AGAMOUS pre-mRNA (Cheng et al. 2003).

In our previous studies, it was found that DNA insertional mutants of *AtCPSF73-II* are lethal (Xu et al. 2004). The hemizygous mutant plants of *AtCPSF73-II* displayed a severe reduction of genetic transmission of female gametes due to a loss of fertility. However, the male gametes were normal. In this report we have demonstrated that knockout/knockdown via RNAi and high over-expression using the CaMV35S promoter result in lethality. In addition, we found that the transgenic plants expressing the *AtCPSF73-I* cDNA under the control of its native promoter caused severe male sterility due to a failure of the anthers to dehisce during flowering (Fig. 7B). Thus, both AtCPSF73-I and -II seem to be associated with flower or reproductive development under the conditions of altered expression level in plants. The impact of both *AtCPSF73*-II mutation and *AtCPSF73-I* expression alteration supports a potential regulatory role in development, probably through interactions with other proteins involved in the pathways. *Arabidopsis* mutants displaying late-dehiscing of anthers have

been identified, and the isolation of the *DELAYED DEHISCENCE1* gene showed that this gene encodes an enzyme in the jasmonic acid synthesis pathway (Sanders et al. 2000). It is possible that the altered expression of *AtCPSF73-I* affects the expression of genes that are involved in jasmonate production potentially through the regulation of the 3′-end formation of these mRNAs. Alternatively, the 3′-end formation of a subset of mRNAs that are closely related to anther dehiscing may be stipulated by extra AtCPSF73-I. Thus, further investigation for interaction partners and the biochemical functions of AtCPSF73-I and -II will help us understand not only the role they play in the cleavage and polyadenylation complex but also their potential role in plant development.

Acknowledgements We thank Dr. Art Hunt for the Gateway vectors of yeast two-hybrid system and clones of AtPAPs and AtCPSF100. We also appreciate the help of Linda Zehler for assisting antibody production, and Cathlene Leary-Elderkin for reading the manuscript. We thank Dr. Detlef Weigel's group for making the microarray data available through their web site. This work was funded by the *Arabidopsis* 2010 Program of NSF (to Q.Q.L.; MCB-0313472), and by grants from Miami University (a postdoctoral fellowship to R.X. from the Center for Functional Genomics and Bioinformatics, a grant from the Committee on Faculty Research to Q.Q.L., and the start-up funds to Q.Q.L. and X.C., respectively).

References

Addepalli B, Meeks LR, Forbes KP, Hunt AG (2004) Novel alternative splicing of mRNAs encoding poly(A) polymerases in *Arabidopsis*. Biochim Biophys Acta 1679:117–128

Aoyama T, Chua NH (1997) A glucocorticoid-mediated transcriptional induction system in transgenic plants. Plant J 11:605–612

Bai C, Tolias PP (1996) Cleavage of RNA hairpins mediated by a developmentally regulated CCCH zinc finger proteins. Mol Cell Biol 16:6661–6667

Bai CY, Tolias PP (1998) Drosophila clipper/CPSF 30 K is a post-transcriptionally regulated nuclear protein that binds RNA containing GC clusters. Nucleic Acids Res 26:1597–1604

Barabino SML, Hübner W, Jenny A, Minvielle-Sebastia L, Keller W (1997) The 30-kD subunit of mammalian cleavage and polyadenylation specificity factor and its yeast homolog are RNA-binding zinc finger proteins. Genes Dev 11:1703–1716

Barabino SML, Ohnacker M, Keller W (2000) Distinct roles of two Yth1p domains in 3′-end cleavage and polyadenylation of yeast pre-mRNAs. EMBO J 19:3778–3787

Belostotsky DA, Meagher RB (1993) Differential organ-specific expression of 3 poly(a)-binding-protein genes from *Arabidopsis thaliana*. Proc Natl Acad Sci USA 90:6686–6690

Bentley D (2002) The mRNA assembly line: transcription and processing machines in the same factory. Curr Opin Cell Biol 14:336–342

Calzado MA, Sancho R, Muñoz E (2004) Human immunodeficiency virus type 1 Tat increases the expression of cleavage and polyadenylation specificity factor 73-kilodalton subunit modulating cellular and viral expression. J Virol 78:6846–6854

Cheng Y, Kato N, Wang W, Li J, Chen X (2003) Two RNA binding proteins, HEN4 and HUA1, act in the processing of AGAMOUS pre-mRNA in *Arabidopsis thaliana*. Dev Cell 4:53–66

Dantonel JC, Murthy KGK, Manley JL, Tora L (1997) Transcription factor TFIID recruits factor CPSF for formation of 3′ end of mRNA. Nature 389:399–402

Dass B, Attaya EN, Wallace A, MacDonald CC (2001) Overexpression of the CstF-64 and CPSF-160 polyadenylation protein messenger RNAs in mouse male germ cells. Biol Reprod 64:1722–1729

Delaney KJ, Xu R, Zhang J, Li QQ, Yun K-Y, Falcone DL, Hunt AG (2006) Calmodulin interacts with and regulates the RNA-binding activity of an arabidopsis polyadenylation factor subunit. Plant Physiol 140(4), in press

Dickson KS, Bilger A, Ballantyne S, Wickens MP (1999) The cleavage and Polyadenylation specificity factor in *Xenopus laevis* oocytes is a cytoplasmic factor involved in regulated polyadenylation. Mol Cell Biol 19:5707–5717

Dinkins RD, Conn HM, Dirk LMA, Williams MA, Houtz RL (2003) The *Arabidopsis thaliana* Peptide Deformylase 1 protein is localized to both mitochondria and chloroplasts. Plant Sci 165:751–758

Dominski Z, Yang XC, Marzluff WF (2005a) The polyadenylation factor CPSF-73 is involved in histone-pre-mRNA processing. Cell 123:37–48

Dominski Z, Yang XC, Purdy M, Wagner EJ, Marzluff WF (2005b) A CPSF-73 homologue is required for cell cycle progression but not cell growth and interacts with a protein having features of CPSF-100. Mol Cell Biol 25:1489–1500

Elliott BJ, Dattaroy T, Meeks-Midkiff LR, Forbes KP, Hunt AG (2003) An interaction between an *Arabidopsis* poly(A) polymerase and a homologue of the 100 kDa subunit of CPSF Plant Mol. Biol 51:373–384

Fields S, Song O (1989) A novel genetic system to detect protein–protein interactions. Nature 340:245–246

Gietz RD, Triggs-Raine B, Robbins A, Graham KC, Woods RA (1997) Identification of proteins that interact with a protein of interest: applications of the yeast two-hybrid system. Mol Cell Biochem 172:67–79

Gutierrez RA, MacIntosh GC, Green PJ (1999) Current perspectives on mRNA stability in plants: multiple levels and mechanisms of control. Trends Plant Sci 4:429–438

Hunt AG, Meeks LR, Forbes KP, Das Gupta J, Mogen BD (2000) Nuclear and chloroplast poly(A) polymerases from plants share a novel biochemical property. Biochem Biophys Res Commun 272:174–181

James P, Halladay J, Craig EA (1996) Genomic libraries and a host strain designed for a highly efficient two-hybrid selection in yeast. Genetics 144:1425–1436

Jenny A, Hauri HP, Keller W (1994) Characterization of cleavage and polyadenylation specificity factor and cloning of its 100-kilodalton subunit. Mol Cell Biol 14:8183–8190

Jenny A, Minvielle-Sebastia L, Preker PJ, Keller W (1996) Sequence similarity between the 73-kilodalton protein of mammalian CPSF and a subunit of yeast Polyadenylation factor I. Science 274:1514–1517

Juge F, Zaessinger S, Temme C, Wahle E, Simonelig M (2002) Control of poly(A) polymerase level is essential to cytoplasmic Polyadenylation and early development in Drosophila. EMBO J 21:6603–6613

Keller W, Nienroth S, Lang KM, Christofori G (1991) Cleavage and Polyadenylation factor CPF specifically interacts with the pre-mRNA 3′ processing signal AAUAAA. EMBO J 10:4241–4249

Kuersten S, Goodwin EB (2003) The power of the 3′UTR: translational control and development. Nat Rev Genet 4:626–637

Kyburz A, Sadowski M, Ditchtl B, Keller W (2003) The role of the yeast cleavage and polyadenylation factor subunit Ydh1p/Cft2p in pre-mRNA 3′-end formation. Nuc Acids Res 31:3936–3945

Li QQ, Hunt AG (1997) The polyadenylation of RNA in plants. Plant Physiol 115:321–326

Li QQ, Das Gupta J, Hunt AG (1998) Polynucleotide phosphorylase is a component of a novel plant poly(A) polymerase. J Biol Chem 273:17539–17543

Li YZ, Chen ZY, Wang WR, Baker CC, Krug RM (2001) The3′-end-processing factor CPSF is required for the splicing of single-intron pre-mRNAs in vivo. RNA 7:920–931

Loke JC, Stahlberg EA, Strenski DG, Haas BJ, Wood PC, Li QQ (2005) Compilation of mRNA polyadenylation signals in *Arabidopsis* revealed a new signal element and potential secondary structures. Plant Physiol 138:1457–1468

Murthy KGK, Manley JL (1992) Characterization of the multisubunit cleavage-polyadenylation specificity factor from Calf thymus. J Biol Chem 267:14804–14811

Ohnacker M, Barabino SM, Preker PJ, Keller W (2000) The WD-repeat protein Pfs2p bridges two essential factors within the yeast pre-mRNA 3′-end processing complex. EMBO J 19:37–47

Proudfoot N (2004) New perspectives on connecting messenger RNA 3′ end formation to transcription. Curr Opin Cell Biol 16:272–276

Ryan K, Murthy KGK, Kaneko S, Manley JL (2002) Requirements of the RNA polymerase II C-terminal domain for reconstituting pre-mRNA 3′ cleavage. Mol Cell Biol 22:1684–1692

Ryan K, Calvo O, Manley JL (2004) Evidence that polyadenylation factor CPSF-73 is the mRNA 3′ processing endonuclease. RNA 10:565–573

Sanders PM, Lee PY, Biesgen C, Boone JD, Beals TP, Weiler EW, Goldberg RB (2000) The *Arabidopsis* DELAYED DEHISCENCE1 gene encodes an enzyme in the jasmonic acid synthesis pathway. Plant Cell 12:1041–1061

Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Schölkopf B, Weigel D, Lohmann JU (2005) A gene expression map of Arabidopsis thaliana development. Nat Genet 37:501–506

Simpson GG, Dijkwel PP, Quesada V, Hemderson I, Dean C (2003) FY is an RNA 3′ end-processing factor that interacts with FCA to control the *Arabidopsis* floral transition. Cell 113:777–787

Takagaki Y, Seipelt RL, Peterson ML, Manley JL (1996) The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B-cell differentiation. Cell 87:941–952

Tarui Y, Minamikawa T (1989) Poly(A) polymerase from *Vigna unguiculata* seedlings - a bifunctional enzyme responsible for both poly(A)-polymerizing and poly(A)-hydrolyzing activities. Eur J Biochem 186:591–596

Vinciguerra P, Stutz F (2004) mRNA export: an assembly line from genes to nuclear pores. Curr Opin Cell Biol 16:285–292

Wesley SV, Helliwell CA, Smith NA, Wang MB, Rouse DT, Liu Q, Gooding PS, Singh SP, Abbott D, Stoutjesdijk PA, Robinson SP, Gleave AP, Green AG, Waterhouse PM (2001) Construct design for efficient, effective and high-throughput gene silencing in plants. Plant J 27:581–590

Xu R, Li QQ (2003) A RING-H2 zinc finger protein gene RIE1 is essential for seed development in *Arabidopsis*. Plant Mol Biol 53:37–50

Xu RQ, Ye X, Li QQ (2004) AtCPSF73-II gene encoding an *Arabidopsis* homolog of CPSF 73 kDa subunit is critical for early embryo development. Gene 324:35–45

Yao YL, Song LH, Katz Y, Galili G (2002) Cloning and characterization of *Arabidopsis* homologues of the animal CstF complex that regulate 3′ mRNA cleavage and polyadenylation. J Exp Bot 53:2277–2278

Zarudnaya M, Kolomiets IM, Hovorun DM (2002) What nuclease cleaves pre-mRNA in the process of Polyadenylation? IUBMB Life 54:27–31

Zhao J, Hyman L, Moore CL (1999) Formation of mRNA 3′ ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. Microbiol Mol Biol Rev 63:405–445